PROTEIN STRUCTURE REPORT

# Crystal structure of the protease-resistant core domain of *Yersinia pestis* virulence factor YopR

FLORIAN D. SCHUBOT, SCOTT CHERRY, BRIAN P. AUSTIN, JOSEPH E. TROPEA, AND DAVID S. WAUGH

Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, Maryland 21702-1201, USA

(RECEIVED March 8, 2005; FINAL REVISION March 25, 2005; ACCEPTED March 28, 2005)

**Abstract**

*Yersinia pestis*, the causative agent of the plague, employs a type III secretion system (T3SS) to secrete and translocate virulence factors into to the cytoplasm of mammalian host cells. One of the secreted virulence factors is YopR. Little is known about the function of YopR other than that it is secreted into the extracellular milieu during the early stages of infection and that it contributes to virulence. Hoping to gain some insight into the function of YopR, we determined the crystal structure of its protease-resistant core domain, which consists of residues 38–149 out of 165 amino acids. The core domain is composed of five α-helices that display unexpected structural similarity with one domain of YopN, a central regulator of type III secretion in *Y. pestis*. This finding raises the possibility that YopR may play a role in the regulation of type III secretion.

**Keywords:** *Yersinia pestis*; plague; type III secretion; YopR; crystal structure

Numerous Gram-negative bacterial pathogens, including *Yersinia pestis*, the causative agent of plague, utilize a type III secretion system (T3SS) to inject effector proteins directly into the cytosol of targeted plant or animal cells (Cornelis 2002). Once the effector proteins have entered the host, they facilitate free proliferation of the bacterium in the infected tissues by suppressing the innate immune response of the host organism. In *Y. pestis*, the structural and secreted components of the T3SS are encoded by a 70-kb virulence plasmid named pCD1.

YopR is probably the least studied of the virulence factors that are secreted via the T3SS in *Y. pestis*. Thus far no enzymatic activity, regulatory role, or other function has been attributed to this 19-kDa protein. In vivo studies examining its role during infection revealed that YopR is not translocated into the cytosol of the host cell but is exported into the extracellular milieu (Lee and Schneewind 1999). The same studies also demonstrated that deletion of the *yopR* gene reduced the virulence of *Y. pestis* 10- to 30-fold in a mouse model of infection. These findings confirmed earlier experiments that had reported a 10-fold increase in the $LD_{50}$ for the *yopR* deletion mutant (Allaoui et al. 1995). Based on these observations, Lee and Schneewind proposed that together with YopB and YopD, two other proteins that are secreted but not delivered into the cytosol of mammalian cells, YopR may be secreted into the extracellular milieu in order to divert host defenses away from the actual site of infection.

Structural studies of YopR were initiated as part of a small-scale structural genomics effort targeting the structural and secreted components of the *Y. pestis* T3SS. We hoped that the structure of YopR would provide clues

Reprint requests to: David S. Waugh, Macromolecular Crystallography Laboratory, Center for Cancer Research, National Cancer Institute at Frederick, P.O. Box B, Frederick, MD 21702-1201, USA; e-mail: waughd@ncifcrf.gov; fax: (301) 846-7148.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051446405.

about its function and role in virulence. Below we describe the effort that culminated in the determination of the three-dimensional structure of the YopR core domain, encompassing amino acids 38–149 (out of 165).

## Results and Discussion

Full-length YopR was readily overproduced in *Escherichia coli* and purified but failed to yield crystals. Therefore, in order to identify an individual structural domain that might be more amenable to crystallization, the sample was subjected to limited proteolysis with thermolysin. This resulted in the identification of a stable digestion product, comprising residues 38–149 ($YopR^{38–149}$), that was ultimately crystallized.

Optimized crystals of $YopR^{38–149}$ diffracted X-rays to a resolution of 1.5 Å. The structure was solved using multiple isomorphous replacement and anomalous scattering (MIRAS) phasing (Table 1). The final model consisted of residues 42–145 of YopR. Residues 38–41, 122–129, and 146–149 appear to have been disordered in the crystal because they were not visible in the electron density maps.

### *YopR resembles the central domain of the* Y. pestis *virulence factor YopN*

The core domain of YopR is composed of five α-helices, four of which are arranged in an antiparallel bundle (Fig. 1A). A DALI search (Holm and Sander 1995) of the Protein Data Bank (PDB) revealed that the closest structural relative of YopR is *Y. pestis* YopN (Schubot et al. 2005), a central regulator of the T3SS that prevents type III secretion from occurring prior to contact with mammalian cells (Forsberg et al. 1991; Day and Plano 1998). YopN consists of three structural domains: an N-terminal chaperone-binding domain, a central α-helical domain of unknown function, and a third domain that mediates its interaction with the coregulatory protein TyeA (Holm and Sander 1995; Iriarte et al. 1998; Schubot et al. 2005). As illustrated in Figure 1A, the four-helix bundle of YopR aligns quite well with the central domain of YopN (Z-score = 7.2, RMSD = 3.6 Å for 71 aligned residues). The significance of this structural homology is unclear, but it is intriguing.

The amino acid sequences of five orthologs of YopR are aligned in Figure 1B. Conserved residues are mainly

**Table 1.** *Data collection and refinement statistics*

(A) Data collection statistics

| | $YopR^{38-149}$-Native | $K_2PtCl_4$ derivative | $K_2HgI_4$ derivative |
|---|---|---|---|
| Resolution [Å] | 54–1.5 | 68–3.2 | 30–2.44 |
| Wavelength [Å] | 1.00 | 1.54 | 1.54 |
| Space Group | $P4_132$ | $P4_132$ | $P4_132$ |
| Cell [Å] | a = 85.62 | a = 86.010 | a = 85.86 |
| Completeness [%] (last shell)[a] | 99.4 (1.55–1.5:97.5) | 99.9 (3.36–3.2:100) | 98 (2.53–2.44:83) |
| Redundancy | 4.5 (3.2) | 11.6 (10.2) | 17 (15) |
| $I/\sigma_I$ | 46.1 (2.3) | 55.3 (12) | 35 (8.7) |
| $R_{merge}$ [%][b] | 3.6 (68) | 5.4 (28) | 9.4 (32) |

(B) Refinement statistics

| | |
|---|---|
| Resolution range [Å] | 54–1.5 |
| R [%] ($I \geq 2\sigma_I$)[c] | 23.9 (20.5) |
| $R_{free}$ [%] ($I \geq 2\sigma_I$)[d] | 27.5 (23.8) |
| Root mean square bonds [Å] | 0.008 |
| Root mean square angles [°] | 1.00 |
| No. of water molecules | 123 |
| Temperature factor [Å$^2$] | 21.6 |
| No. of protein atoms | 1562 |
| No. of solvent molecules | 123 |
| Ramachandran analysis (%) | |
| Most favored | 96.3 |
| Allowed | 3.7 |
| Disallowed | 0 |

[a] The values in parentheses relate to the highest resolution shell.
[b] $R_{merge} = \Sigma|I - <I>|/\Sigma I$, where I is the observed intensity, and $<I>$ is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.
[c] $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively.
[d] $R_{free}$ defined in Brunger (1992).

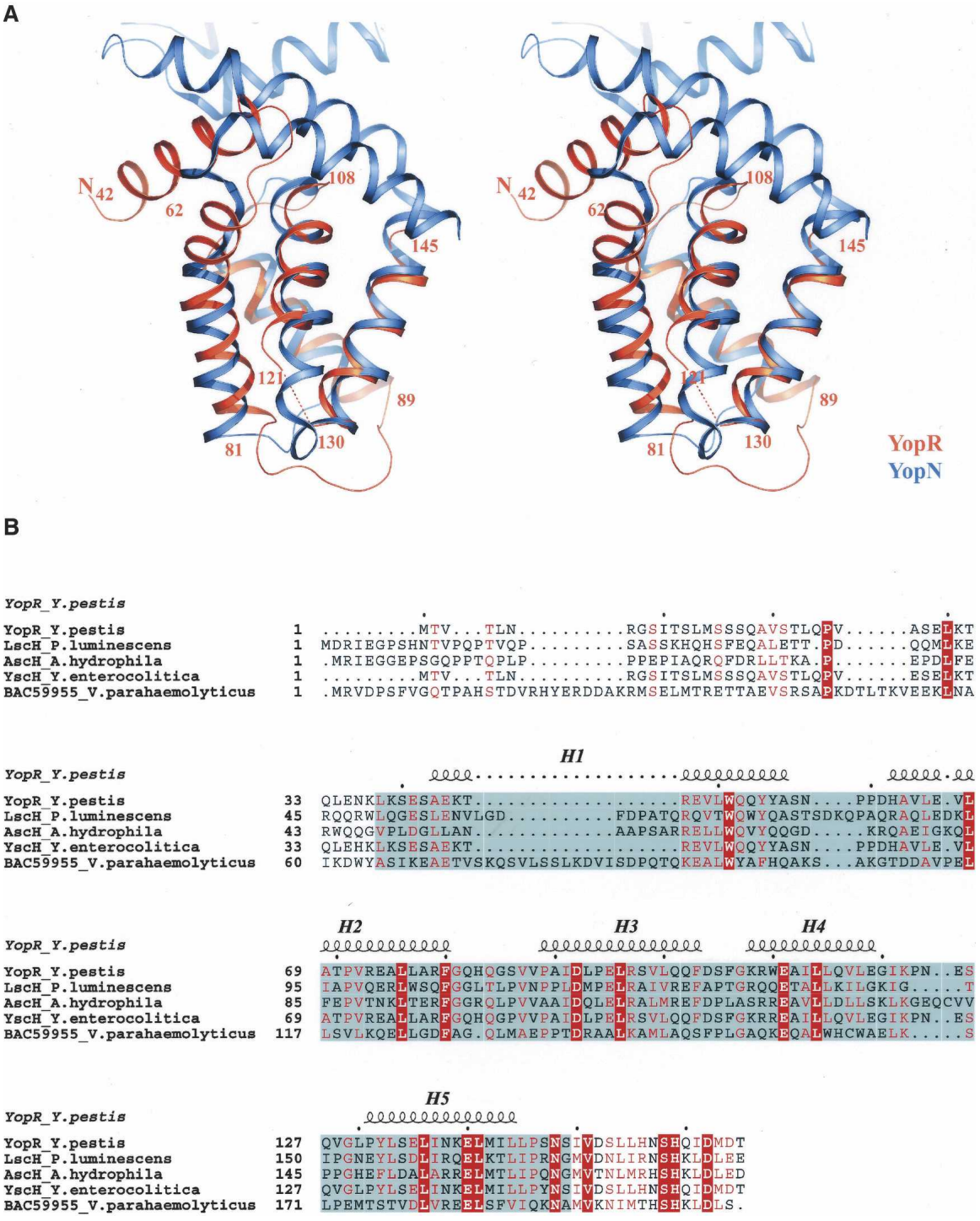


**Figure 1.** (*A*) PyMOL-generated (DeLano Scientific) superposition of the YopR$^{38–149}$ structure and the central domain of YopN encompassing residues 78–165. (*B*) ESPript-generated (Gouet et al. 1999) sequence alignment of YopR orthologs. Shaded gray is the region corresponding to the crystallized fragment.

concentrated near their C termini. Since the C terminus of YopR is particularly well conserved, an extra effort was made to determine the structure of a YopR construct that contained the final 15 residues of the protein (YopR$^{38–165}$). Although we were able to overproduce and purify YopR$^{38–165}$ (data not shown),

unfortunately we were unable to crystallize it. Residues 150–158 are predicted to form an additional α-helix (Rost et al. 2004). An examination of the electrostatic surface properties of YopR revealed a large hydrophobic cavity on its surface (data not shown) that might accommodate the missing C-terminal helix. The bond that was cleaved by thermolysin (between residues 149 and 150) is predicted to be located in a loop region between helix H5 and the missing C-terminal helix. If this conjecture is correct, then we can conclude that the C-terminal helix is not essential for the structural integrity of the core domain of YopR.

In conclusion, although there is a noteworthy similarity between the structure of the central domain of YopN and YopR, the structure of the latter molecule did not improve our understanding of what its function may be. One possibility is that, like YopN, YopR plays a role in the regulation of type III secretion. In any case, the availability of the structure of YopR should facilitate future efforts to investigate its role in virulence.

## Materials and methods

### *Expression and purification of YopR-His$_6$*

An expression vector encoding full-length YopR was assembled by Gateway recombinational cloning. A recognition site for tobacco etch virus (TEV) protease and a hexahistidine tag were added to the N and C termini of YopR, respectively, during PCR. The PCR amplicon, also flanked by appropriate *att* recombination sites, was inserted into pDONR201 (Invitrogen) to generate the entry clone pKM956. The DNA sequence of the insert was verified and subsequently recombined into the destination vector pKM596 (Fox and Waugh 2003) to create the expression vector pKM964. This vector was designed to produce YopR in the form of an "affinity sandwich" with maltose-binding protein (MBP) and a hexahistidine tag joined to its N and C terminus, respectively.

In order to achieve in vivo cleavage of the fusion protein by TEV protease, single colonies of *E. coli* BL21(DE3) CodonPlus RIL cells (Stratagene) containing pRK603, a TEV protease expression vector (Kapust and Waugh 2000), and pKM964 were grown to saturation in 200 mL of Luria broth supplemented with 100 μg/mL ampicillin, 30 μg/mL chloramphenicol, and 30 μg/mL kanamycin at 37°C. The saturated culture was diluted 1:50 into 6 L of the same medium and grown to early log phase ($A_{600} = 0.3$–$0.5$) at 37°C, at which point the temperature was shifted to 30°C (the optimum temperature for TEV protease processing), and both isopropyl-B-D-thiogalactopyranoside (IPTG) and anhydrotetracycline (final concentrations of 1 mM and 100 ng/mL, respectively) were added to initiate the production of the fusion protein and TEV protease. Four hours later, the cells were recovered by centrifugation at 5000*g* for 15 min.

Thirty-five grams of cell paste were resuspended in 350 mL of 50 mM sodium phosphate buffer (pH 8.0), 300 mM NaCl, and 25 mM imidazole (buffer A) along with four tablets of Complete EDTA-free protease inhibitor cocktail (Roche Molecular Biochemicals). The cells were lysed with an APV Gaulin Model G1000 homogenizer at 10,000 psi and centrifuged at 23,000*g* for 30 min at 4°C. The supernatant was filtered through a 0.45-μm polyethersulfone membrane and then applied to a 25-mL Ni-NTA Superflow affinity column (Qiagen) equilibrated with buffer A. The column was washed with five column volumes of buffer A. The sample was eluted using a 10-column volume gradient to 100% buffer B (200 mM imidazole, 50 mM sodium phosphate [pH 8.0], and 300 mM NaCl).

The relevant fractions were pooled and loaded onto a sephacryl S-100 column (Amersham Biosciences) equilibrated with 25 mM Hepes (pH 7.0) and 100 mM NaCl (buffer C). The peak fractions containing the YopR were pooled and concentrated to 12 mg/mL. The final product was judged to be >95% pure by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (data not shown). The molecular weight was confirmed by electrospray mass spectrometry. Aliquots were flash-frozen with liquid nitrogen and stored at −80°C until use.

### *Limited proteolysis and purification of YopR$^{38–149}$*

A 1 mg/mL stock solution of thermolysin (Roche Molecular Biochemicals) in water was used for the limited proteolysis experiments. The YopR-His$_6$ stock solution consisted of the protein at 1 mg/mL in buffer C. The five individual reactions were composed of 25 μL of YopR-His$_6$ stock solution, 25 μL of 2× thermolysin buffer (20 mM Tris-HCl [pH 8.0], 4 mM $CaCl_2$, 0.4 M NaCl, and 10% glycerol), and 0.5 μL of serial 1:4 dilutions of the thermolysin stock solution. The reactions were allowed to proceed for 1 h at 37°C, at which time they were stopped by the addition of 0.5 μL of 0.5 M EDTA. The reaction products were initially analyzed by SDS-PAGE. The precise molecular weights of the fragments were obtained by electrospray mass spectrometry.

The large-scale thermolysin digest of YopR-His$_6$ was performed by combining 5 mL of 5 mg/mL YopR-His$_6$, 5 mL of 2× thermolysin buffer, and 0.1 mL of thermolysin at 0.25 mg/mL. The reaction proceeded for 1 h at 37°C and was stopped by the addition of 0.1 mL of 0.5 M EDTA. The sample was then loaded onto a Sephacryl S-100 column equilibrated with buffer C. The peak fractions were concentrated to 30 mg/mL. The final product was evaluated by SDS-PAGE and electrospray mass spectrometry. Aliquots were flash-frozen with liquid nitrogen and stored at −80°C until use.

### *Crystallization of YopR$^{38–149}$*

Crystallization screening of the truncated sample was conducted in Vapor Batch Plates (Hampton Research) using the modified microbatch technique (Chayen 1997) in which 1 μL protein and 1 μL crystallization screening solution are mixed and covered with 2 mL of a 50:50 mixture of paraffin and silicone oil (Hampton Research). The sample was initially screened with commercially available crystallization matrices, and crystals were obtained from condition 31 of the WIZARD I crystallization Screen (Emerald BioSystems). The refined conditions, consisting of 17% PEG-8000, 85 mM phosphate-citrate (pH 4.2), 0.17 M NaCl, and 15% glycerol, yielded the cubic crystals used for the structure solution.

### *X-ray data collection*

Crystals of $YopR^{38–149}$ were mounted in a loop without additional cryo-soaking and subsequently flash-frozen in liquid nitrogen. To gain phase information, $K_2PtCl_4$ and $K_2HgI_4$ derivatives were generated by soaking $YopR^{38–149}$ crystals for 1 d in the crystallization buffer solution containing 1 mM $K_2PtCl_4$ and 1 mM $K_2HgI_4$, respectively. The native data set used for the structure solution was collected on a Brandeis CCD detector at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) beamline X9B. Data sets for the two heavy atom derivatives were collected using a MAR-345 image plate mounted on a Rigaku X-ray generator. Data processing was carried out with the HKL2000 program suite (Otwinowski and Minor 1997). The details of data collection and processing for all data sets are provided in Table 1.

### *Structure solution and refinement*

The native and derivative data sets were analyzed in SOLVE (Terwilliger and Berendzen 1999), and the resulting MIRAS phases were directly channeled into RESOLVE (Terwilliger 2000). The initial 2.0 Å map and partial backbone trace created by RESOLVE were of excellent quality, exhibiting clear protein/solvent boundaries and recognizable features of protein secondary structure. After density modification, nearly the entire backbone and most of the side chains could be traced by ARP/wARP (Perrakis et al. 1999). The structure was manually completed with the molecular modeling program O (Jones et al. 1991). The model was refined with REFMAC (Murshudov et al. 1997) followed by manual adjustment against SIGMAA (Collaborative Computational Project 4 1994) weighted difference Fourier maps. After several rounds of manual adjustment and refinement, 123 water molecules were added to the structure using ARP/wARP in combination with REFMAC.

Model quality was assessed with PROCHECK (Laskowski et al. 1993). All nonglycine residues in the structure resided either in the most favorable or in the allowed regions of the Ramachandran plot, and the overall geometry was better than average when compared with structures solved at the same resolution. Model refinement statistics are given in Table 1. The atomic coordinates and structure factors for the $YopR^{38–149}$ structure have been deposited in the PDB (Berman et al. 2000) with accession code 1Z21.

## Acknowledgments

Electrospray mass spectrometry experiments were conducted on the LC/ESMS instrument maintained by the Biophysics Resource in the Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute at Frederick. Some X-ray diffraction data were collected at the National Synchrotron Light Source X9B beamline.

## References

Allaoui, A., Schulte, R., and Cornelis, G.R. 1995. Mutational analysis of the *Yersinia enterocolitica* virC operon: Characterization of yscE, F, G, I, J, K required for Yop secretion and yscH encoding YopR. *Mol. Microbiol.* **18:** 343–355.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28:** 235–242.

Brunger, A.T. 1992. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355:** 472–475.

Chayen, N.E. 1997. The role of oil in macromolecular crystallization. *Structure* **5:** 1269–1274.

Collaborative Computational Project 4. 1994. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50:** 760–763.

Cornelis, G.R. 2002. The *Yersinia* Ysc-Yop "type III" weaponry. *Nat. Rev. Mol. Cell. Biol.* **3:** 742–752.

Day, J.B. and Plano, G.V. 1998. A complex composed of SycN and YscB functions as a specific chaperone for YopN in *Yersinia pestis*. *Mol. Microbiol.* **30:** 777–788.

Fox, J.D. and Waugh, D.S. 2003. Maltose-binding protein as a solubility enhancer. *Methods. Mol. Biol.* **205:** 99–117.

Forsberg, A., Viitanen, A.M., Skurnik, M., and Wolf-Watz, H. 1991. The surface-located YopN protein is involved in calcium signal transduction in *Yersinia* pseudotuberculosis. *Mol. Microbiol.* **5:** 977–986.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. 1999. ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15:** 305–308.

Holm, L. and Sander, C. 1995. Dali: A network tool for protein structure comparison. *Trends Biochem. Sci.* **20:** 478–480.

Iriarte, M., Sory, M.P., Boland, A., Boyd, A.P., Mills, S.D., Lambermont, I., and Cornelis, G.R. 1998. TyeA, a protein involved in control of Yop release and in translocation of *Yersinia* Yop effectors. *EMBO J.* **17:** 1907–1918.

Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47:** 110–119.

Kapust, R.B. and Waugh, D.S. 2000. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expr. Purif.* **19:** 312–318.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26:** 283–291.

Lee, V.T. and Schneewind, O. 1999. Type III machines of pathogenic yersiniae secrete virulence factors into the extracellular milieu. *Mol. Microbiol.* **31:** 1619–1629.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53:** 240–255.

Otwinowski, Z.M. and Minor, W. 1997. HKL2000. *Methods Enzymol.* **276:** 307–326.

Perrakis, A., Morris, R., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6:** 458–463.

Rost, B., Yachdav, G., and Liu, J. 2004. The PredictProtein server. *Nucleic Acids Res.* **32:** W321–W326.

Schubot, F.D., Jackson, M.W., Penrose, K.J., Cherry, S., Tropea, J.E., Plano, G.V., and Waugh, D.S. 2005. Three-dimensional structure of a macromolecular assembly that regulates type III secretion in *Yersinia pestis*. *J. Mol. Biol.* **346:** 1147–1161.

Terwilliger, T.C. 2000. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56(Pt. 8):** 965–972.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol Crystallogr.* **55(Pt. 4):** 849–861.